

DNA sequences of group 4 allergens from rye, wheat, barley and *Lolium perenne*

Comparison with isoforms of *Phleum pratense* Phl p 4

A. Nandy, M. Wald, L. Gräfe, O. Cromwell, H. Fiebig

Allergopharma Joachim Ganzer KG, R&D Department, 21465 Reinbek, Germany

Contact e-mail: andreas.nandy@allergopharma.de

Introduction

Grass pollen allergy is one of the most important allergic diseases world-wide. Several grass species grown in meadows, like *P. pratense* and *L. perenne*, contribute to allergic sensitisations, but also allergens from extensively cultured cereals, especially rye, make a profound contribution to the development of allergy. The group 4 major allergen of *P. pratense*, Phl p 4, is recognised by more than 70 % of grass allergic patients^{1,2,3}. IgE-binding cross-reactivity has been described for some group 4 allergens of different grass species³, but until now only the Phl p 4 gene could be deciphered on the DNA-level.

Results

The Poaceae group 4 allergens represent a family of basic proteins with molecular weights of about 55 kDa and calculated pI values far above 8 (Tab. 1, Fig. 1). In rye, wheat and *P. pratense* distinct isoforms with amino acid identities of 88 to 94 % could be detected. Additionally these isoforms exist in different minor variants. The inter-species homology lies in the range 83 % (Phl p 4 to *Trifolium* species) to 95 % (Sec c 4 to Tri a 4) (Fig. 2, Fig. 3).

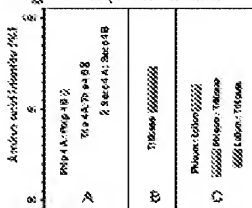
Tab. 1 Sequence analysis of grass pollen group 4 allergens

Protein	Source	Sequence length (amino acids)	Isoelectric point (pI)	Molecular weight (Da)
Phl p 4 A	<i>Phleum pratense</i>	509	8.8	55,895
Phl p 4 B	<i>Phleum pratense</i>	509	8.2	55,924
Lol p 4*	<i>Lolium perenne</i>	423 (fragment)	8.8*	*
Sec c 4 A	<i>Secale cereale</i> (rye)	498	8.1	54,939
Sec c 4 B	<i>Secale cereale</i> (rye)	498	8.3	54,953
Tri a 4 A	<i>Triticum aestivum</i> (wheat)	497	8.9	55,257
Tri a 4 B	<i>Triticum aestivum</i> (wheat)	497	8.9	55,149
Hor v 4	<i>Hordeum vulgare</i> (barley)	498	8.5	54,815

The sequence length, isoelectric point and molecular weight calculations were made on the basis of the mature proteins. For Phl p 4 the N-terminal residue has been determined by N-terminal protein sequencing. Most of the homology alignment (Fig. 1) the relative divergence rates of *Trifolium* spp. have been used for calculation.

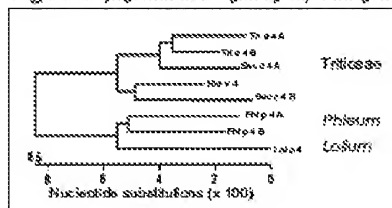
* The Lol p 4 sequence is only partial and contains about 85 % of the mature Lol p 4 sequence.

Fig. 2 Sequence identities



The amino acid identities were calculated on the basis of the mature allergens. In case of Lol p 4 the overlapping region has been used for calculation. A: The sequence identities of conserved variants of group 4 allergens range from 85 % (Lol p 4) to 95 % (Tri a 4). The two major variants of Phl p 4 show inter-species identities of 82 %. B: The identities of allergens of the *Trifolium* species range from 82 % (Sec c 4 B) to Tri a 4 A to 95 % (Sec c 4 A) to Tri a 4 B. C: Inter-species identities of members of the genera *Phleum*, *Lolium* and of the *Trifolium* species *Medicago*, *Trifolium* species.

Fig. 3 Phylogenetic tree of grass group 4 allergens

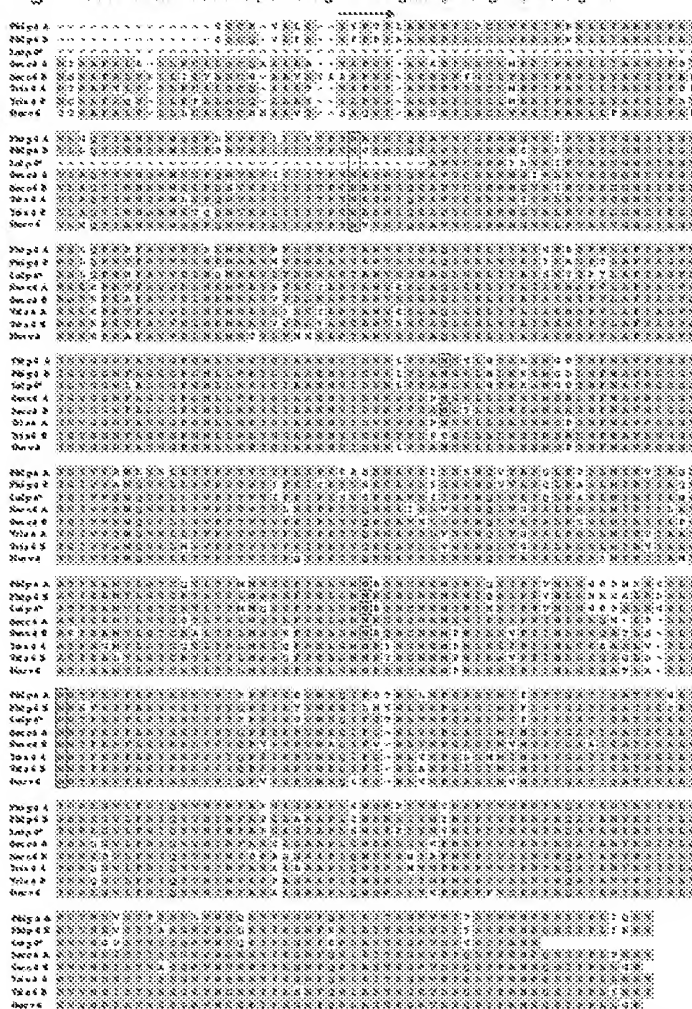


The dendrogram illustrates the phylogenetic relationships of the grass group 4 allergens. The rooted tree has been generated by using the DNA sequence data that overlap the Lol p 4 fragment (127 bp). Remarkably, inter-species variants (e.g. Sec c 4 A and B, Tri a 4 A and B) show sequence identities similar to those of sequences originating from different *Trifolium* species (compare also Fig. 2A and B). The amino acid identities for *Phleum* species variants that have similar degrees of amino acid difference as compared to the *Lolium* perenne sequence (compare also Fig. 2A and B).

Methods

Based on the DNA sequence of Phl p 4 several PCR-primer sequences with cross-reactivity to DNA sequences of related species could be designed. The group 4 DNA sequences of *Lolium perenne* (Lol p 4), *Secale cereale* (Sec c 4), *Hordeum vulgare* (Hor v 4), and *Triticum aestivum* (Tri a 4) have been amplified, cloned and sequenced.

Fig. 1 Deduced amino acid sequence alignment of grass pollen group 4 allergens



Multiple alignment of *Phleum pratense* Phl p 4 variant forms, *Lolium perenne* Lol p 4, *Secale cereale* (rye) Sec c 4 variant forms (which Tri a 4 variant forms, and *Hordeum vulgare* (barley) Hor v 4. Residues that match the consensus sequence, the *Phleum* sequence. The start of the mature Phl p 4 sequence as deduced by N-terminal protein sequencing of purified mature Phl p 4 is marked with an arrow. Potential N-glycosylation sites are marked with blue boxes.

* The Lol p 4 sequence is only partial and contains about 85 % of the mature Lol p 4 sequence.

Conclusion

The group 4 allergens represent a family of proteins that are conserved among different grass species. The occurrence of cross-reacting isoforms in distinct species with amino acid homologies that are comparable to those of different group 4 molecules across the species border is remarkable. Since recombinant group 4 allergens may be important for a future

References

- 1) R.E. Rossi et al. (2001). *Allergy* 56, 1160-11853
- 2) K. Andersson and J. Lidholm (2003). *Int. Arch. Allergy Immunol.*, Review article, 130, 67-107
- 3) S. Stumvoll et al. (2002). *Biol. Chem.* 383, 1363-135
- 4) LaserGene DNASTAR, Inc., Madison, WI 53715, U.S.A.